Parallel Design the Algorithm of Digital Images Segmentation Split & Merge

Andrea Reyes
Biomedical Engineering Research Group
UIS, Colombia
andrea.reyes@ieee.org

Víctor Martínez
Biomedical Engineering Research Group
UIS, Colombia
viedumar@uis.edu.co

Ana Ramírez
Connectivity and Signal Processing Group
UIS, Colombia
anaberam@uis.edu.co

Abstract

This work presents the design of a parallel algorithm based in the Split & Merge regions method (Horowitz and Pavlidis, 1974) for the Segmentation process of Digital Images, in this case Uterine-Cervix samples.

The proposal utilizes the methodology for the designing parallel algorithm of Ian Foster (Partitioning, Communication, Agglomeration and Mapping).

1. Introduction

Image Segmentation is a key step in Image Processing. Subdivide an image into constituent objects, called regions; it is the process of assigning pixels to each region having common properties, using image attributes such as pixel intensity, spectral values and textural properties [3, 5, 15].

Let R represent the entire image having different objects. An Segmentation of R is a partition of R into subsets $R_1, R_2, \ldots, R_n$, such that [12]:

(a) $U_{i=1}^{n} R_i = R,$

(b) R_i *is a connected region,* $\forall i,$

(c) $R_i \cap R_j = \emptyset \ \forall i,j, i \neq j,$

(d) $P(R_i) = TRUE \ \forall i,$

(e) $P(R_i \cup R_j) = FALSE \ \forall i,j, i \neq j,$

Where $P(R_i)$ is a logical predicate over the set of pixels in the set of pixels in R_i and $\emptyset$ is the empty set.

Proposition (a) indicates that segmentation must be complete, that is, every pixel must be in a region while the second one indicates the pixels belonging to a region must be connected. The proposition (c) represents that the regions must be disjointed, and (d) deals with the properties that must be satisfied by the pixels in every segmented region R_i in such a way $P(R_i)$ = TRUE if all pixels in R_i are equivalent with respect R_i and R_j are different in the sense of predicate P.

The Segmentation algorithms can be based in two approaches [5]: First, the discontinuities of the pixels, which the main areas of interest are the detection isolated points and the detection of edges and lines. Second, the similarity of the pixels, where is usual to use algorithms such as threshold or oriented regions (Growth or Split & Merge), since grouped the pixels according to similar properties. The algorithms used for the detection of limits and threshold are not very effective, due to they do not make use of spatial information, so the oriented regions are most suitable for the type of Uterine-Cervix medical images.

The region growing algorithm uses a set of points generators from which it will increases the region, the fundamental problem is to correctly determine the set of initial points. Another alternative is the algorithm Split & Merge, it subdivides the original image into a set of arbitrary disjunct regions and then merge the regions to try to satisfy the criteria of homogeneity, steps that solves the region growing problem. On the other hand, the method based on regions, to determine the location of the final frontier, it has the difficulty of a high consumption of computational resources, which is particularly relevant in applications that require real time response and for this reason, the parallel algorithms constitute a very important approach to solve this problem.

The Split & Merge Algorithm proposed by Horowitz and Pavlidis [2, 4] in 1974 is one of the most popular algorithms for image segmentation.

The algorithm may be summarized by the following steps [3, 8]:

(1) Split any region R_i into four almost equal regions where $P(R_i)$ = FALSE.
(2) Merge any adjacent regions R_i and R_j for which $P(R_i \cup R_j)$ = TRUE.
(3) Stop when no further merging or splitting is possible.
Otherwise repeat steps (1) and (2).

The application area is the Medical Images for this research and the input images are Uterine-Cervix samples. The Cervical Cancer is the second type of cancer most frequently in the female population; the diagnostic is that near 80% of the cases in not developed countries. The test of Cervical Cancer cytology is used to establish the population of women at risk of this cancer, in Bucaramanga (Colombia) finding 1.897 cases of slight displace, 736 moderate displaces, 289 several displaces, in the years 2000 and 2001 [13, 18].

In the Biomedical Engineering Research Group (GIIB) the Industrial of Santander University (UIS), has been performed research works in undergraduates and postgraduates in the field of Digital Image Processing in the detection of Cervical Cancer, however there have been problems in the stage of segmentation, where the main inconvenient to implement an image processing system to support analysis, is that the images acquired in the test are large size and quantity, and the time is quite large [10, 11, 13, 14, 16, 17].

Faced with this situation it is proposed to design a parallel algorithm based on the regions method Split & Merge, with the purposed to posterior implement and evaluate the reduction in the execution time, since at present has not been achieved an optimal sequential implementation, due to that in different tests performed has been come to use all the computational resources available without to obtain a result.

2. Split & Merge Regions Method

The Algorithm Split & Merge requires two types of operations [3]; a fast split phase which is followed by one or more merge phases. The split stage partitions an image into square regions which conform to a first homogeneity criterion; then merge these square regions into larger regions which conform to a second homogeneity criterion.

The Split stage consists of generating a quad-tree (see Fig. 1), where nodes at level i are the result of splitting the sub- images of size $N_1 / 2^i$ x $N_2 / 2^i$ into four sub-images of size $N_1 / 2^{(i+1)}$ x $N_2 / 2^{(i+1)}$ [3, 5]. The construction of this quad-tree can be carried out following a Bottom-up strategy of merging sub-images or Top-Down strategy of splitting [2, 4].

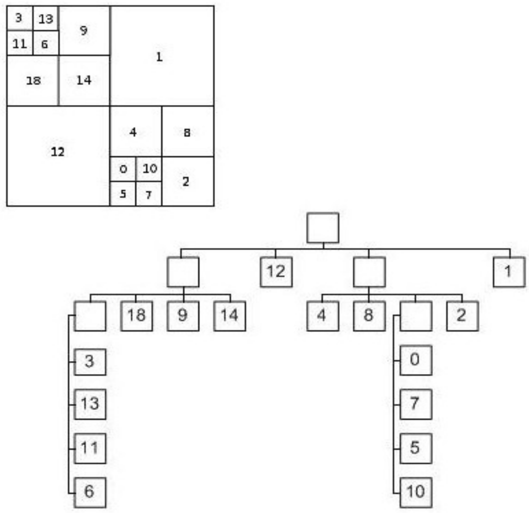

Figure 1. Quad-tree representation

In both strategies, only those nodes satisfying the homogeneity criterion will be generated. The Split phase ends when no more square regions can be generated. The set of leaves (regions) on the quad-tree produced by the Split phase will be the set of input regions for the Merge phase.

The homogeneity criterion of the one region is given by the gray levels of pixels that comprise, establishing a criterion of uniformity. Let $X_{i,\ i=1,2,3,\ldots n}$ the gray levels of pixels that make up a region R. The average value of gray levels in the region R, is the arithmetic average of gray levels of pixels [9]:

$$\bar{X}_R = \frac{\sum_{n}^{i=1} X_i}{n} \qquad (1)$$

A region R is uniform gray levels, if for every sub-regions $r_i \in R$, is satisfied that the ratio between the minimum of the average values of gray levels of the R region and a sub-region r_i and the maximum of these values is less than or equal to 0.95. That is,

$$\frac{Min\left(\bar{X}_R, \bar{X}_{ri}\right)}{Max\left(\bar{X}_R, \bar{X}_{ri}\right)} \leq 0{,}95 \ \forall \, ri, i = 1, \ldots 4 \qquad (2)$$

And the region is homogeneity in gray levels if it satisfies with the criterion of uniformity.

3. Design Parallel Algorithm of Ian Foster

This methodology structures the design process as four following stages (see Fig. 2). In the first and two stages, the focus is on concurrency and scalability. In the third and fourth stages, attention shifts to locality and performance [1].

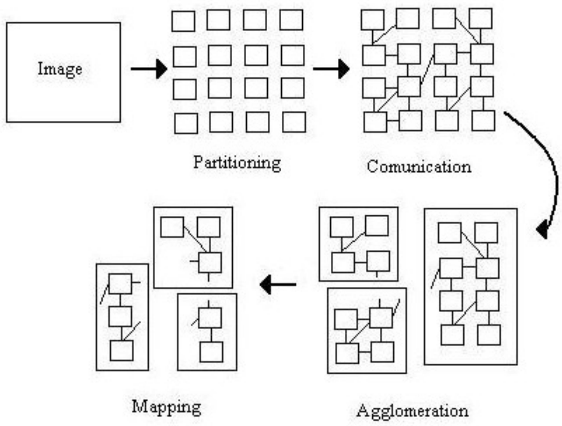

Figure 2. Methodology for design parallel programs

i) Partitioning: The partitioning stage of a design expose opportunities for parallelism. The focus is on defining a large number of small tasks, fine-grained decomposition of a problem.

ii) Communication: The tasks require data associated with another task. Data must then be transferred between tasks so as to allow computation to proceed. In the communication is defining a channel structure that links and specifies the messages that are to be sent and received on these channels. The communication can be classified in local-global, structured-unstructured, static-dynamic, and synchronous-asynchronous.

iii) Agglomeration: The task and communication structures determined in the two previous stages are evaluated against the requirements of performance and costs implementation, to obtaining an algorithm that will execute efficiently on some class of parallel computer. If necessary, the tasks are combined to increase performance and reduce the communication and may still be greater than the number of processors.

iv) Mapping: In the final stage, specify where each task is to execute and assigned to each processor. Operating system or hardware mechanisms can be relied upon to schedule executable tasks to available processors. The goal in developing mapping algorithms is minimize total execution time.

4. Proposed Algorithm

i) Partitioning: The Split step is inherently parallel task. Initially, an image of size $N = N_1 \times N_2$ is partitioned into $P = P_1 \times P_2$ sub-images that containing adjacent pixels $N_1/P_1 \times N_2/P_2$ [6]. The structure for represent the regions of the image is the Quad-tree [5, 7].

ii) Communication: Each processor work splitting and merging regions, this corresponds to either removing or building nodes of the tree structure [5], communicates transferred data of the adjacency node, this information is stored using an adjacency matrix. Each processor communicates too with others processors for merge regions, satisfying the homogeneity criterion, and the constraints (a)-(e) in the steps (1)-(3) (see section 1).

iii) Agglomeration: In the quad-tree representation are defined the method for identifying regions (Id), for this method use a cyclic and random strategy [3, 6], each processor assigns an Id to each region. The Ids should be different for each region and regions in every processor. The Id assigned in the quad-tree contains the information the vertices and edge the region. For select the task concurrently use a load balancing describes in the next step.

iv) Mapping: The process of the merge between regions are data dependent and the performance of communication, agglomerate and map the algorithm decreasing with the execution time, in this case the algorithms for static and dynamic load balancing are necessaries. For the load balance locally the dynamic balanced algorithm, assigning identifiers cyclic, is suitable and the dynamic balanced algorithm with assignment identifiers random for the load balance global [3].

The parallel algorithm can be describes by the following steps [3, 6, 5, 7, 8]:

Step I: Partitioning the image P1 x P2 in four sub-images and mapped onto the set of Processors.
Step II: Each Processor realized the Split on its own sub-image and assigns an identifier (Id) to each of the created regions.
Step III: Each Processor builds its own local tree by creating the set of edges between its own local regions which satisfy homogeneity criteria.
Step IV: Processors exchange information about the regions at the border of the initial sub-image and create the set of edges between regions that belong to different neighboring Processor.
Step V: Each region determines the linked regions that best suit the homogeneity criterion as regions to be merged. If more than one region are candidates to

merge with it will be selected that region with the minimum Id.

Step VI: Processors must exchange information about the best selection for those linked regions located at different Processor.

Step VII: The vertices and edges of the tree are updated for the current set of regions.

Step VIII: Run the algorithm for load balancing.

Step IX: Repeat steps V–VIII while linked regions still exist.

Otherwise the algorithm ends.

5. Discussions and Future Work

The algorithm proposed is for implement in a SPMD programming model and working in architecture MIMD with the message passing MPI.

The importance of this research is given to contribute to the development of the line of research on Digital Image Processing in the area of High Performance Computing on the GIIB; in the solution to the problem of the cost of time processing images Uterine-Cervix for diagnostic Cervical Cancer.

6. References

[1] Foster, I. "Designing and Building Parallel Programs: Concepts and Tools for Parallel Software Engineering". *Addison-Wesley Longman Publishing Co*., Inc. 1995.

[2] Horowitz, S.L. and Pavlidis, T. "Picture segmentation by a directed split-and-merge procedure". *Proceedings, 2nd International J. C. on Pattern Recognition, Copenhagen,* 424–433. 1974

[3] Montoya, M. D. Gil, C. and García, I. "The load unbalancing problem for region growing image segmentation algorithms". *J. Parallel Distrib. Comput.* 63, 4 (Apr. 2003), 387-395. 2003.

[4] Horowitz, S. L. and Pavlidis, T. "Picture Segmentation by a Tree Traversal Algorithm". *J. ACM 23, 2 (Apr. 1976*), 368-388. 1976.

[5] Kelkar, D. and Gupta, S. "Improved Quadtree Method for Split Merge Image Segmentation". I*n: Proceedings, First international Conference on Emerging Trends in Engineering and Technology - Volume 00 (July 16 - 18, 2008). ICETET. IEEE Computer Society, Washington, DC,* 44-47. 2008.

[6] Montoya, M.D.G.; Gil, C.; Garcia, I., "Load balancing for a class of irregular and dynamic problems: region growing image segmentation algorithms," *Parallel and Distributed Processing. Proceedings of the Seventh Euromicro Workshop on* , vol., no., pp.163-169, 3-5, Feb 1999.

[7] Tyagi, A.; Bayoumi, M., "Systolic array implementation of image segmentation by a directed split and merge procedure," *Pattern Recognition, 1990. Proceedings., 10th International Conference on* , vol.ii, no., pp.491-493 vol.2, 16-21, Jun 1990.

[8] Faruquzzaman, A.B.M.; Paiker, N.R.; Arafat, J.; Karim, Z.; Ameer Ali, M., "Object segmentation based on split and merge algorithm," *TENCON 2008 - 2008. IEEE Region 10 Conference* , vol., no., pp.1-4, 19-21, Nov. 2008.

[9] Alvarez M, Rivas M. and Rukoz M. "Segmentación de Imágenes Biomédicas Mediante el Crecimiento de Regiones". *Acta Científica Venezolana,* 52: 192–198, 2001

[10] Martinez, V. et al. 2007. "Tecnología Grid para la Detección de Cáncer de Mama y Cuello Uterino por Medio de Procesamiento de Imágenes". *CLCAR 2007. Colombia, 13-18 Agosto* 2007, 318-324.

[11] Niño, D. et al. "Contribución al estudio de las células escamosas de citologías cérvico uterinas que presentan cambio por ASCUS, por medio de tratamiento digital de imágenes". *Trabajo de Grado. UIS. 2006.*

[12] Gonzalez, R.C. Woods, R.E. "Digital Image Processing", *2nd Edition, Prentice-Hall, Inc., Upper Saddle River, NJ*. 2002.

[13] Martinez, V. et al. "Computational model for squamous cells characterization during cervical smear cytology" . *Rev. Colomb. Biotecnol. Vol. VII No. 2 Diciembre* 2005, 35-46. 2005.

[14] Amaya, M.L. et al. "Contribución al estudio de las Células Escamosas de Citologías Cérvico Uterinas que Presentan Cambios por Virus de Papilona Humano (VPH), utilizando Tratamiento Digital de Imágenes". *Trabajo de Grado. UIS.* 2006.

[15] Tilton, J.C., "Image segmentation by iterative parallel region growing with applications to data compression and image analysis," *Frontiers of Massively Parallel Computation. Proceedings., 2nd Symposium on the Frontiers of* , vol., no., pp.357-360, 10-12 Oct 1988

[16] Martinez, V. "Modelo Computacional para Caracterización de células endocervicales". *Trabajo de Maestria. UI*S. 2007.

[17] Martinez, V.E. et al. "Software Tool for Squamous Cells Classification in Cervical Smear Cytologies" *ICCB2005: Proceedings of II International Conference on Computational Bioengineering, IST Press, Lisbon*, 2:1089-1094. 2005.

[18] Universidad Autónoma de Bucaramanga. 2002. "Registro de población de cáncer en el área metropolitana de Bucaramanga". 2000-2001.